Volume 214, number 2, 226–235 FEB 04591 April 1987

*Discussion Letter*

# A model of the nucleotide-binding site in tubulin

Himan Sternlicht, Michael B. Yaffe and George W. Farr

*Department of Pharmacology, Case Western Reserve University, Cleveland, OH 44106, USA*

Received 19 January 1987

Tubulin uses GTP to regulate microtubule assembly and is thought to be a member of a class of GDP/GTP-binding proteins (G-proteins) as defined by Hughes [(1983) Febs Lett. 164, 1–8]. How tubulin is structurally related to G-proteins is not known. We use a synthesis of sequence comparisons between tubulin, other G-proteins, and ADP/ATP-binding proteins and topological arguments to identify potential regions involved in nucleotide binding. We propose that the nucleotide-binding domain in the $\beta$-subunit of tubulin is an $\alpha/\beta$ structure derived from amino acid residues ~60–300. Five peptide sequences are identified which we suggest exist as 'loops' that extend from $\beta$-strands and connect $\alpha$-helices in this structure. We argue that GDP binds to four of the five loops in an $Mg^{2+}$-*independent* manner while GTP binds in an $Mg^{2+}$-*dependent* manner to a different combination of four loops. We propose that this switch between loops upon GTP binding induces a conformational change essential for microtubule assembly.

G-protein; Microtubule assembly; Protein structure

## 1. INTRODUCTION

Tubulin, the major constituent protein of microtubules, is a heterodimer consisting of two homologous polypeptide chains, $\alpha$- and $\beta$-tubulin, each ~450 amino acid residues in length [1–4]. The fact that microtubules are involved in a diverse variety of cellular processes suggests that this protein has a large number of binding sites for macromolecules and ligands, such as MAPs, colchicine and GTP. It has been proposed that tubulin is folded into a structure with functionally distinct binding domains or regions formed from spatially discrete sequences [5–7]. How these regions are organized in relation to each other in tubulin is not known.

Tubulin binds two molecules of GTP tightly: one exchangeably, at the so-called E-site in $\beta$-tubulin [8,9], and one non-exchangeably, presumably at a site in $\alpha$-tubulin. One $Mg^{2+}$ is also bound tightly, either indirectly as a nucleotide-metal ion complex [10] or possibly directly to certain residues in the protein [11]. Tubulin's conformation is nucleotide-dependent and differs if GTP or GDP is bound at the E-site [12–14]. These conformational differences are thought to be biologically important (review [15]). GTP, in contrast to GDP, is a potent effector of microtubule assembly when bound at the E-site. In addition, microtubule ends containing tubulin-GTP are more stable than those containing tubulin-GDP. This stability difference is thought to result in the coexistence of mixed populations of growing and shrinking microtubules at 'steady state' (referred to as 'dynamic instability') [16].

In 1983 Hughes proposed that guanine nucleotide-binding proteins including tubulin define a

Correspondence address: H. Sternlicht, Dept of Pharmacology, Case Western Reserve University, Cleveland, OH 44106, USA

*Abbreviations:* MAPs, microtubule-associated proteins; $\alpha/\beta$, a type of fold observed in proteins consisting of $\beta$-strands and their interconnecting $\alpha$-helices organized as a sheet of predominantly parallel $\beta$-strands with the $\alpha$-helices packed against the front and back faces of the sheet [40]



*Published by Elsevier Science Publishers B.V. (Biomedical Division)*
00145793/87/$3.50 

distinct class of regulatory proteins referred to as G-proteins [17]. He noted that (i) tubulin, like other G-proteins, exerts biological function through protein-protein interactions modulated by nucleotide binding; (ii) binding of GTP induces a conformation change in tubulin which facilitates self-association into microtubules; and (iii) GTP hydrolysis which accompanies microtubule assembly leaves GDP non-covalently bound to tubulin subunits in the microtubule and facilitates depolymerization.

Despite the striking similarities between tubulin and the G-proteins, recent efforts to identify the nucleotide-binding site in tubulin on the basis of sequence homology with other G-proteins and thus confirm Hughes' hypothesis were largely unsuccessful [18,19]. In this paper we use a synthesis of sequence comparisons between tubulin, other G-proteins, and the ADP/ATP-binding proteins [3,7] and topological arguments [20] to generate a model for the GTP/GDP-binding site of $\beta$-tubulin. In our model the nucleotide-binding domain contains ~45–55% of the amino acid residues of $\beta$-tubulin and is folded into an $\alpha/\beta$ structure. We suggest that, except for an additional phosphoryl-binding loop, the nucleotide binding site in $\beta$-tubulin may be similar to the nucleotide-binding site observed in EF-Tu and proposed for the Ras protein p21 [19,21,22]. Because of this additional binding loop GDP binding in $\beta$-tubulin is $Mg^{2+}$-independent whereas GTP binding is $Mg^{2+}$-dependent. As a result, GTP induces conformational changes in $\beta$-tubulin by a different mechanism from that in EF-Tu.

## 2. SEQUENCE COMPARISONS BETWEEN TUBULIN AND OTHER G-PROTEINS

The only three-dimensional structure of a GTP-binding protein known to date is EF-Tu complexed to GDP. In this complex the nucleotide-binding site was shown to consist of four loops which con-

Table 1

Sequence homologies of the GTP/GDP-binding site in G-proteins[a]

| | Phosphoryl-binding regions | | Guanine-binding region | Spacing between regions | |
|---|---|---|---|---|---|
| | I | II | III | I–II | II–III |
| Putative consensus sequences | G X X X X G K | D X X G | N K X D | | |
| Elongation and initiation factors | | | | | |
| EF-Tu, *E. coli* | G H V D H G K | D C P G | N K C D | 56 | 52 |
| EF-1 alpha, yeast, human, *A. salina* | | A | M | 71 | 59 |
| IF-2, *E. coli* | | T | I | 40 | 51 |
| EF-G, *E. coli* | A I A | T | M | 65 | 51 |
| Ras proteins | | | | | |
| ras-1, yeast | G G G G V G K | D T A G | N K S D | 41 | 56 |
| ras-2, yeast | | | L | 41 | 56 |
| H-, N-, K-ras, human | A | | C | 41 | 56 |
| Signal-transducing proteins | | | | | |
| Transducin $\alpha$, $\gamma$ | G A G E S G K | D V G G | N K K D | 154 | 66 |
| $G_s$ $\alpha$, rat brain | | | Q | 170 | 66 |
| $G_i$ $\alpha$, rat brain | | | | 155 | 66 |

[a] A select list from table 1 of Dever et al. [23]

In each class of proteins only differences in sequences are indicated. Single-letter codes are used to designate amino acid residues

nect a central core of $\beta$-strands to $\alpha$-helices [19]. Similar binding sites are thought to exist in other GTP-binding proteins [21,22]. Dever et al. [23] noted from their survey of the amino acid sequences of more than 25 GTP-binding proteins, that regions homologous to three of EF-Tu's nucleotide-binding loops could be identified. These homologous regions, referred to as regions I, II and III in table 1, occur in a similar order and spacing to that observed for loops I–III of EF-Tu and were presumed to be components of the GTP/GDP nucleotide-binding sites in these proteins [23]. The spacing between regions I and II ranges from 40 to 75 residues while that between regions II to III ranges from 51 to 68 residues (table 1). These values contrast well with the spacings of 56 and 52 residues, respectively, observed for loops I and II, and II and III in EF-Tu [19]. In transducin and adenylate cyclase $G_s$ and $G_i$, however, the spacing between regions I and II ranges from 154 to 170 residues and is significantly larger than that observed for EF-Tu, while the spacing between regions II and III is constant at 66 residues, similar to that observed for EF-Tu (table 1). As a result, these proteins may correspond to a unique subclass of G-proteins [23].

The putative consensus sequences listed in table 1 were derived by analyzing homologous regions I–III for invariant residues [23]. Invariant residues in the consensus sequences presumably correspond to essential residues that interact specifically with guanine nucleotides. X-ray analysis of EF-Tu–GDP reveals that GDP is in the *anti* conformation with its $\alpha$-phosphate bound to loop I via

Table 2

Proposed sequences for the GTP/GDP-binding site in tubulin are conserved

| | | | | |
|---|---|---|---|---|
| Putative consensus sequence from table 1 | GXXXXGK | (GXXGXG)* | DXXG | NKXD |
| Regions of tubulin sequence homology | I (105–112) | $I_A$ (143–148) | II (205–208) | III (297–300) |
| $\beta$-Tubulins[a–d] | $_{111}$G E T Y H G K $_{105}$[a–d] | $_{143}$G G T G S G $_{148}$[a,b,d] (A[c]) | $_{205}$D N E A $_{208}$[a–d] | $_{300}$N K A D $_{297}$[a] (S[b] P[c] N[d]) |
| $\alpha$-Tubulins[e,f] | $_{106}$G H Y T I (V[f]) G K $_{112}$[e] | $_{143}$G G T G S G $_{148}$[c] | $_{205}$D N E A $_{208}$[c,f] | N.F. |

* Tubulins appear to have an additional conserved region $I_A$ not generally found in G-proteins but seen in ATP-binding proteins (see text). Single-letter codes are used to designate amino acid residues. Numbers represent amino acid positions when $\alpha$- and $\beta$-tubulin are aligned for maximal sequence homology (cf. Mandelkow et al. [7]). Note that regions I and III in $\beta$-tubulin are found in the opposite (C- to N-terminal) orientation relative to the putative consensus sequences. When amino acid sequences differ from tubulins referenced in a and e, the differences are displayed. (a) Nine vertebrate tubulins [Sullivan and Cleveland (1986) Proc. Natl. Acad. Sci. USA 83, 4327]; one mouse tubulin (M $\beta$ 5) [Lewis et al. (1985) J. Cell Biol. 101, 852]; two *Chlamydomonas* tubulins (*C. reinhardtii*) [Youngbloom et al. (1984) Mol. Cell Biol. 4, 2686]; one *Trypanosome* tubulin (*T. brucei rhodesiense*) [Kimmel et al. (1985) Gene 35, 237]; one yeast tubulin (*S. cerevisiae*) [Neff et al. (1983) Cell 32, 211]; one sea urchin tubulin, incomplete fragment [Alexandraki and Ruderman (1983) J. Mol. Evol. 19, 397]. (b) Four vertebrate tubulins [Sullivan and Cleveland (1986) Proc. Natl. Acad. Sci. USA 83, 4327]. (c) One *Neurospora* tubulin (*N. crassa*) [Orbach et al. (1986) Mol. Cell Biol. 6, 2452]. (d) One yeast tubulin (*S. pombe*) [Hiraoka et al. (1984) Cell 39, 349]. (e) Two human tubulins [Cowin et al. (1983) Mol. Cell Biol. 3, 1738]; porcine brain tubulin (Krauhs et al. (1981) Proc. Natl. Acad. Sci. USA 78, 4156]; one chicken tubulin [Valenzuela et al. (1981) Nature 289, 650]; one rat tubulin [Lemishka and Sharp (1982) Nature 300, 330]; two mouse tubulins [Lewis et al. (1985) J. Cell Biol. 101, 852]; one *Trypanosome* tubulin (*T. rhodesiense*) [Kimmel et al. (1985) Gene 35, 237]; two *Chlamydomonas* tubulins (*C. reinhardtii* [Silflow and Youngbloom (1986) Ann. NY Acad. Sci. 466, 18]. (f) Two yeast tubulins (*S. pombe* [Toda et al. (1986) Cell 37, 233]; one *Physarum* tubulin (*P. polycephalum*) [Krammer et al. (1985) J. Mol. Biol. 183, 633]. N.F., not found in these proteins

the invariant lysine residue, and its $\beta$-phosphate bound via *an* $Mg^{2+}$ to the invariant aspartic acid residue in loop II. The guanine ring in EF-Tu-GDP is bound in a pocket formed from a hydrophobic loop (loop IV), which is not conserved among the G-proteins [23], and loop III, which provides guanine specificity via hydrogen bonds to the invariant aspartate and asparagine residues in that loop. Further structural details of EF-Tu-GDP are given in the paper by La Cour et al. [19].

Dever et al. estimated the odds of a chance occurrence of the three consensus sequences ordered as in table 1 with spacings 40–80 residues to be 0.01–0.02%, in a protein containing 1000 residues [23]. Furthermore, when they searched a protein data base (Protein Identification Resource, NIH) which contained ~3800 sequences for the simultaneous occurrence of the three consensus sequences, with no restrictions placed on sequence spacing, they found only two additional proteins not in their original compilation. One was subsequently shown not to bind GTP while the second was predicted to be a GTP-binding protein [23]. If they further allowed one conservative amino acid replacement to occur in the consensus sequences (A for G, E for D and Q for N) they then found six additional proteins. However, only one of the six had a realistic spacing ($\leq$200 amino acid residues between sequences I and III, and $\geq$20 amino acid residues between II and III). It is not known whether this protein binds GTP ([23]; also Merrick, W., personal communication).

We searched the published sequences of tubulin, which presently consists of ~20 $\beta$-tubulin and 13 $\alpha$-tubulin sequences from a large variety of species and tissues, for conserved sequences that conformed to the consensus sequences in table 1, or to the consensus sequences with alanine residues replacing glycine residues (this conservative substitution, for example, has been observed in EF-G, cf. table 1). The results are summarized in table 2. Since $\alpha$- and $\beta$-tubulins have high sequence homology (>40%) and both subunits are thought to bind guanine nucleotides, we initially scored a consensus sequence in $\alpha$- or $\beta$-tubulin as positive only if the region is conserved in both subunits at essentially identical residue positions. Only two regions in $\alpha$- and $\beta$-tubulin met these initial conditions. These regions, designated respectively as regions I and II in table 2, occur at residue positions 105–112 and 205–208 and have the consensus sequences GXXXXGK and DXXG/A characteristic of phosphoryl binding. When we relaxed the requirement that the putative consensus sequence must be present in both subunits at similar positions, we found the third conserved region (table 2). Region III is composed of residues 297–300 in $\beta$-tubulins, and has the consensus sequence NKXD characteristic of guanine-binding specificity. Additional regions in $\beta$-tubulin having the consensus sequence DXXG/A were also found but these occurred at the extreme C-terminus or N-terminus ends of the $\beta$-tubulin sequence and were rejected as improbable (see below). The spacing between regions I and II in the $\beta$-tubulins is 93 residues while that between regions II and III is 89 residues. These values are not much different from those observed for the G-proteins shown in table 1. Thus, both the sequential order and the spacing between the conserved regions argue that $\beta$-tubulins may have a guanine nucleotide-binding site similar to that observed in EF-Tu. Following Dever et al. [23] we estimated the odds that regions I, II and III, spaced at least 50–100 residues apart, would occur by chance in a 450-residue-long protein such as $\beta$-tubulin to be $\leq$0.1%*. This conservative estimate included G/A substitutions in region II and permitted regions I and III to be in either orientation.

In table 2 we also included an additional conserved region which occurs at residue positions 143–148 in all tubulins but which is not generally found in other G-proteins. This glycine-rich

* Following Dever et al. [23] it was assumed that: (i) all amino acids occur with equal probability (1/20); (ii) the spacing between consensus sequences GXXGK, DXXA and NKXD is 50–100 amino acids; (iii) sequences I and III can occur in either the forward or reverse direction, and an A as well as a G is acceptable in II. The approximate probability of the three sequences occurring in a 450 amino acid length protein, such as $\beta$-tubulin, is therefore:

$$\sim(2)(1/20)^3(300)(2)(1/20)^2(50)(2)(1/20)^3(50) = 2 \times 10^{-4} \text{ or } 1/5000$$

While the assumptions may not be entirely valid (for example, G and K often exceed 5%), it would appear that the chance occurrence of these three sequences as specified above is $\leq$0.1%.

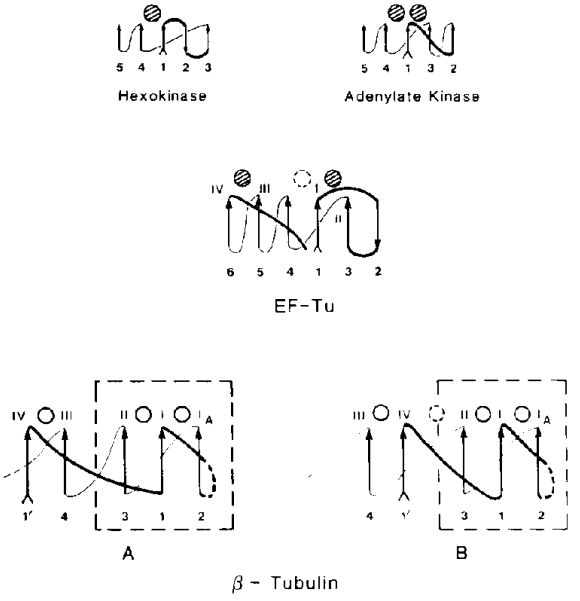


Fig.1. Proposed connectivity diagrams for $\beta$-tubulin. The topology of the proposed phosphoryl-binding core in $\beta$-tubulin (dashed enclosures in A and B), or the core modified to include a hydrophobic loop and a guanine-specific loop (A, B), is compared with the topologies of the GTP/GDP-binding protein EF-Tu and the ATP/ADP-binding proteins adenylate kinase [43] and hexokinase (large domain) [45]. Adenylate kinase has a glycine-rich phosphoryl-binding loop, $I_A'$ at the C-terminus end of $\beta$-strand 1 [43] which is homologous to region $I_A$ in tubulin. On the other hand, hexokinase, which lacks the glycine-rich phosphoryl-binding sequence [44], differs topologically from adenylate kinase and binds ATP differently [45]. $\beta$-strands (dark arrows) are numbered sequentially in accordance with their location in the primary sequence, starting with strand 1 or 1′ as the most N-terminal. Arrowheads denote C-terminal ends of the $\beta$-strands. A connection, which consists of $\alpha$-helices and turns, emerges first from the C-terminus of a $\beta$-strand (i.e. strand $n$) and then proceeds to the N-terminus end of the next strand in the sequences (strand $n+1$). Connections which pack against the front face of the $\beta$-sheet are indicated by thick lines while connections which pack against the back face of the sheet are indicated by thin lines. Unknown or uncommitted folds are indicated by heavy dashed lines. Potential substrate-binding clefts can be identified from strand order in $\alpha/\beta$ proteins (see text). Circles depicted with full-line circumferences denote predicted sites which have amino acid sequences compatible with ligand binding, while dashed-line circumference circles denote topologically predicted sites that lack appropriate binding sequences and consequently are less probable. Hatched circles denote that binding of substrate was observed in the predicted region (cf. [20,21]). Taking adenylate kinase as an example, strand order initially proceeds left to right from strand 1 to 2, then reverses at strand 2 and proceeds right to left. As a result, connection '1 to 2' packs on the front face of the $\beta$-sheet, while connections '2 to 3', '3 to 4', and '4 to 5' pack on the back face. Strand 1 is thus adjacent to two strands, 3 and 4, whose connections are on the opposite side of the sheet relative to strand 1, and two binding clefts are consequently predicted. These predicted binding clefts have been observed with nucleotide analogs. Further details as well as a discussion of hexokinase can be found in Branden's review [20]. In A and B of $\beta$-tubulin, Branden's rules led to the prediction of three and four binding sites, respectively. However, the topologically identified site formed from loops II and IV (dashed circumference circle) is an overprediction as loop II in our model is committed to phosphoryl binding while loop IV lacks phosphoryl-binding sequences (see text).

region, which we denote as $I_A$ in table 2, has been previously implicated as a possible phosphoryl binding site in tubulin [7,24]. It is homologous to the glycine-rich phosphoryl-binding loop found in lactate dehydrogenase and conforms to the consensus sequence GXXGXG involved in phosphoryl binding in a variety of ATP/ADP-binding proteins [25,26]. Furthermore, the separation between regions $I_A$ and II is 57 residues, well within the range of values observed for phosphoryl-binding loops in other G-proteins. We estimate the probability for sequences I, $I_A$, II and III, spaced 50–100 residues apart, to have occurred by chance to be less than 0.01%. Curiously, regions I, $I_A$ and II are clustered in a relatively small portion (~20%) of the sequence of tubulin, with region $I_A$ between regions I and II. We believe that this clustering of regions, each apparently capable of phosphoryl binding, is not coincidental (estimated probability of chance occurrence $<0.1\%$). Rather, we suspect that all three regions participate in phosphoryl binding.

## 3. TOPOLOGICAL CONSIDERATIONS

In order to explore the spatial relationships between regions I–III in $\beta$-tubulin we use a topological approach modelled after Branden [20]. The nucleotide-binding domain in tubulin is likely

to be an $\alpha/\beta$ structure, an assumption consistent with circular dichroism studies of calf brain tubulin which indicate ~26% $\alpha$-helix and ~47% $\beta$-sheet at physiological pH and room temperature [27]. In $\alpha/\beta$ proteins nucleotide-binding sites or clefts are generally formed from loops that extend from the carboxy ends of the parallel strands of the $\beta$-sheet. In 1980 Branden showed that the location of these clefts can be predicted from the order of the $\beta$-strands within the sheet. Fig.1, for example, shows schematic diagrams for the nucleotide binding domain in hexokinase (large domain), adenylate kinase and EF-Tu as well as our models for $\beta$-tubulin. In these diagrams sheet twist is ignored and the $\beta$-sheet is taken as a planar structure. The $\beta$-strands of the sheet are interconnected by $\alpha$-helices and turns. Sheets formed from parallel strands, for example, have connections that leave from the C-terminus end of the $\beta$-sheet to crossover and re-enter the sheet from the N-terminus end. Since these crossover connections are right-handed*, when the order of the $\beta$-strands within the sheet reverses the connection switches from one face of the sheet to the other (fig.1). When two adjacent strands in the interior of the sheet have connections on opposite faces, conditions are geometrically favorable for the formation of clefts [20]. The loops which extend from these adjacent strands form ligand- and co-factor-binding sites at the C-terminal end of the sheet (fig.1).

In EF-Tu, for example, GDP is bound primarily at the carboxy-terminus ends of a $\beta$-sheet with the phosphoryl moiety in one cleft and the guanine ring in a second one [19]. This sheet consists of 5 parallel strands and 1 anti-parallel strand (fig.1).

* Terminology according to Richardson [41]. Consider 2 adjacent parallel strands 1 and 2 in a $\beta$-sheet. If these strands are ordered left to right as '1 followed by 2' when viewed from the top of the sheet, the cross-over connection from 1 to 2 will occur on the front face of the sheet for right-handed connections, and on the back face for left-handed connections. If the strands are ordered left to right as '2 followed by 1' when viewed from the top, the cross-over connection from 1 to 2 will occur on the back face of the sheet for right-handed connections, and on the front face for left-handed connections. Left-handed connections appear to be rare [41].

Focusing on the parallel strands, strand 1 is adjacent to strands 3 and 4, whose connections are on the opposite side of the sheet relative to strand 1 (cf. adenylate kinase). Consequently, this strand order reversal predicts two binding clefts, but only one is observed (fig.1). The observed binding site is formed from the phosphoryl-binding loops I and II which extend from the C-terminus ends of strands 1 and 3, respectively. A third binding cleft is predicted at the C-terminus ends of strands 5 and 6 at the edge of the $\beta$-sheet. Even though strand 6 is the last strand in the sheet and does not physically connect to another strand, the polypeptide segment which leaves its C-terminal end to cross over the front face of the sheet to emerge near the N-terminus end of strand 1 is topologically equivalent to a true connection from strand 6 to strand 1. The predicted binding cleft is observed, and is formed from loops III and IV which provide guanine-binding specificity [19].

Branden's rules often overpredict the number of binding sites since they pertain to fold topology and not to the actual binding potentials of the amino acids in the ligand-binding loops. However, because these rules do not appear to underpredict the number of sites [20], they are a powerful method for identifying potential binding sites. In the case of EF-Tu, for example, Branden's rules predict three nucleotide-binding clefts, but only two of these have loops that contain amino acid residues capable of interacting with GTP (i.e. consensus sequences I–III). Consequently, in developing our model we 'tempered' these rules by requiring each binding site component to contain both GTP-binding sequences and the correct fold topology to generate a binding cleft.

We assumed that the GTP-binding site of $\beta$-tubulin is composed of two components: a phosphoryl-binding component and a guanine-binding component. Sequences I, $I_A$ and II of $\beta$-tubulin (table 2) are homologous to those implicated in GTP/GDP- and ATP/ADP-binding proteins. We began with the working assumption that all three sequences function as phosphoryl-binding sites in $\beta$-tubulin, and constructed a model for the phosphoryl-binding region. Sequences I, $I_A$ and II were taken as loops at the C-terminus ends of $\beta$-strands and ordered as shown in fig.1A and B (dashed enclosures). This fold topology predicts two phosphoryl-binding clefts: one cleft at the C-

Table 3

Homologous hydrophobic regions in $\alpha$- and $\beta$-tubulin

| | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | $^{64}$RAVFVDL | $^{111}$GKEIIDL | $^{147}$SGFTSLLM | $^{229}$RLISQIVSS | $^{267}$FPLATVAPV | $^{312}$YMACCLLYR |
| $\beta$ | $^{64}$RAILVDL | $^{111}$GAELVDS | $^{147}$SGMGTLLI | $^{229}$HLVSATMS | $^{267}$FFMPGFAPL | $^{312}$YLTVAAVFR |

Hydrophobic regions in human $\alpha$ [42] and chicken $\beta$ [4] tubulin were located by means of the Kyte and Doolittle algorithm [29] – a procedure which assigns hydropathy constants to the residues based on their hydrophobicities and that of their neighbors. In our study hydropathy constants were calculated with the aid of a program written in Forth for a Commodore 64 PC computer (unpublished) using a running average over a 9-residue window [29]. Positive values indicative of hydrophobic regions were obtained at ~120 residues in $\alpha$- and ~100 residues in $\beta$-tubulin. Homologous hydrophobic regions, i.e. hydrophobic regions occurring at the same sequence position in $\alpha$- and $\beta$-tubulin, involved ~50 residues in $\alpha$- and $\beta$-tubulin, or ~50% of the hydrophobic residues (a value comparable to the overall value of ~45% sequence homology between the two subunits). The homologous regions are listed above

terminus ends of strands 1 and 2 involving loops I and $I_A$, and the second cleft at the C-terminus ends of strands 1 and 3 involving loops I and II. Aside from a mirror image equivalent arrangement this represents *the only way strands 1, 2 and 3, with their corresponding loops I, $I_A$ and II, can be ordered to generate two predicted binding clefts by Branden's rules.* Interestingly, residues in loops I and $I_A$ bind phosphates in an $Mg^{2+}$-independent manner, whereas residues in loop II interact with the phosphoryl moiety in an $Mg^{2+}$-dependent manner [19,25]. Hence, by appropriate placement of the phosphoryl moiety this model can rationalize the paradoxical observation that GTP binding to tubulin is strongly dependent on magnesium concentration, whereas GDP binding is essentially independent of magnesium [11,28] (see section 4).

Continuing with our model, we assume that the guanine binding component of $\beta$-tubulin is formed from two loops: a conserved loop (loop III) providing guanine specificity in all known GTP-binding proteins [23], and a hydrophobic loop (loop IV) which is not conserved among other GTP-binding proteins. We attempted to locate loop IV in $\alpha$- and $\beta$-tubulin based on results of a previous study which used the Kyte and Doolittle algorithm [29,30]. A total of six homologous hydrophobic regions present in both $\alpha$- and $\beta$-tubulin were identified (table 3). We dismissed the three hydrophobic regions that fell within our proposed phosphoryl-binding 'core' (residues 90–230). The hydrophobic region 312–330 was also dismissed because this region is located too close to loop III (residues 297–300) to permit an $\alpha$-helix–$\beta$-strand–turn geometry. This left only two hydrophobic regions: one from residues 64–70, and a second region from residues 265–275. UV cross-linking studies performed with 8-azido GTP (Kim and Haley, personal communication) are consistent with residues 64–70 as the hydrophobic loop, and the analysis was consequently continued with this region. Residues 60–69 of $\beta$-tubulin were also independently implicated by Leberman and Egner [18] as a potential site for guanine base binding although their basis for selection differs from ours.

The complete model of the GTP-binding site, including the phosphoryl-binding core together with the guanine-binding cleft, is shown in fig.1A and B. For simplicity the nucleotide-binding domain is depicted as containing five $\beta$-strands, the minimum number required to generate three binding clefts. In this model we depict loop IV as emerging from strand 1′ and loop III as emerging from strand 4. Strand 1′ was placed adjacent to strand 4 and has a connection crossing over the front-face of the $\beta$-sheet to strand 1. In fig.1A strand 1′ is to the left of strand 4. In fig.1B strand 1′ is to the right of strand 4. These two structures represent the only ways strands 1′–4, with their corresponding loops I–IV, can be ordered by Branden's rules to generate one guanine-binding site and two phosphoryl-binding sites.

## 4. DISCUSSION

In this study we explored the relationship between $\beta$-tubulin which binds the exchangeable

GTP important for microtubule assembly and GTP-binding proteins. To the best of our knowledge, this study is the first to combine a topological approach with sequence comparisons to construct a model of a nucleotide-binding domain in a cytoskeletal protein. Conventional methods for predicting protein structure focus on $\alpha$-helices and $\beta$-strands as folding elements in proteins and tend to ignore random coil or loop structures [31]. In contrast, this study, which is based on Branden's rules for ligand binding in $\alpha/\beta$ proteins [20], focuses on the arrangement of loops at the carboxy-terminus ends of parallel $\beta$-strands and their connecting helices. Our model consists of two components. One portion of the GTP/GDP-binding site was formed from sequences I, $I_A$ and II, homologous to phosphoryl binding sequences found in a variety of GTP/GDP- and ATP/ADP-binding proteins. These three sequences displayed as loops in fig.1A and B can be uniquely ordered by Branden's rules to generate two predicted phosphoryl-binding clefts. Experimental evidence supports this two-phosphoryl-binding site model for $\beta$-tubulin (see below). The second portion of the binding site is composed of loops III and IV which confer guanine-binding specificity. We noted that the requirement that loops I–IV form a guanine-specific binding cleft and two phosphoryl-binding clefts constrained the fold topology by Branden's rules to 2 possible structures (fig.1A,B). One structure has loop IV to the left of loop III. The second structure has loop IV to the right of loop III. Further studies are required to distinguish between our two equally plausible structures. Although we developed these models independently of direct comparisons with any known protein structures, fig.1A has an especially close topological correspondence to the nucleotide-binding domain in EF-Tu.

We do not known at present whether our model is accurate although various observations indicate that it is reasonable:

(i) Secondary structure predictions support the proposed model (fig.1B). For example, we predict using the Garnier et al. [32] algorithm (Glynias, M., Yaffe, M.B. and Sternlicht, H., unpublished) [30] that region I in $\alpha$- and $\beta$-tubulin is part of a turn or random coil which subsequently develops into an amphiphilic helix; region $I_A$ is in a turn in $\alpha$- and $\beta$-tubulin, and region II is part of a turn and helix in $\beta$-tubulin. Similarly, Little et al. [24] using the Chou and Fasman [33] algorithm report that residues 132–160 in $\alpha$- and $\beta$-tubulin, which contain region $I_A$, is likely to have a $\beta$-strand, loop and $\alpha$-helix structure, at positions 132–139, 140–152 and 153–160, respectively.

(ii) In EF-Tu, two phosphoryl-binding sites are predicted by Branden's rules. However, sequences implicated in GTP binding are present only at the site formed from loops I and II (fig.1). Loop II has been shown to bind the $Mg^{2+}$ of the GDP-Mg complex. This $Mg^{2+}$ appears to be essential for the binding of both GTP and GDP to EF-Tu [34]. The situation for $\beta$-tubulin, however, is somewhat different. Correia et al. [28] found that tubulin formed a ~1400-fold tighter complex with GTP-Mg than with GTP alone. In contrast, tubulin's affinity for GDP was unaffected by the presence of a magnesium ligand and, within a factor of 2–3 was the same as that for the GTP-Mg complex. Our topological model has sequences capable of phosphoryl binding at each of the binding sites predicted by Branden's rules. Based on the amino acid sequence, nucleotide binding to one of these sites via loops I and $I_A$ should be $Mg^{2+}$-independent, whereas binding to loops I and II present at the other site should be $Mg^{2+}$-dependent. This model can then explain the results of Correia et al. [28] as follows: $\alpha$- and $\beta$-phosphate groups of GDP and GDP-Mg could bind to loops I and $I_A$ in an $Mg^{2+}$-independent manner, as observed [11,28], and the $\beta$- and $\gamma$-phosphates of GTP, which presumably are sterically constrained from binding to loop $I_A$, could bind instead to loops I and II when complexed to $Mg^{2+}$. (In the absence of magnesium, GTP may bind only weakly to this site, due to electrostatic repulsion between the invariant aspartate residue in loop II and the $\beta$- and $\gamma$-phosphate groups). We postulate that the switch from loop $I_A$ to II with GTP-Mg binding would trigger a conformational change which induces microtubule assembly. Conversely, GTP hydrolysis causes a switch from loop II back to $I_A$ and would trigger a conformational change which favors disassembly. This conformational change mechanism is different from that proposed to be operative in EF-Tu which lacks region $I_A$ [19,21,22].

(iii) Our model predicts that the presence of

$Mg^{2+}$ is essential for nucleotide binding to loops I and II, thereby inducing a conformational change for microtubule assembly. This explains the long-standing observation that microtubule assembly is dependent on $Mg^{2+}$ (cf. [35]). The model can also explain the recent observation by Bayley and Manser [36] that pyrophosphate-Mg is an effector of assembly. According to our model, the absence of the guanine ring and the presence of $Mg^{2+}$ allow the pyrophosphate to bind to loops I and II, rather than to loops I and $I_A$. Thus, our model predicts that in the absence of magnesium, pyrophosphate alone will not support microtubule assembly. This prediction awaits experimental verification.

(iv) Nath et al. [37] demonstrated that the guanine base cross-links to portions of $\beta$-tubulin both proximal and distal to Tyr 281, consistent with our proposed loop IV (residues 64–70) and loop III (residues 297–300). Linse and Mandelkow [46] showed that GTP cross-links to the first 90 residues at the N-terminus end of $\beta$-tubulin. Studies done with 8-azido GTP under stringent binding conditions demonstrate a single major cross-link to $\beta$-tubulin which occurs in region 64–70 (Kim and Haley, personal communication).

(v) Sullivan and Cleveland [38] noted that amino acid substitutions in vertebrate $\beta$-tubulins cluster at residue positions 30–57 and 420–450, demonstrating evolutionary conservation of our proposed GTP-binding domain.

In EF-Tu and presumably in $\beta$-tubulin, guanine nucleotides bind in the *anti* conformation. In EF-Tu–GDP ribose hydroxyls are directed away from the protein [19]. Our topological model (fig.1B) predicts that when nucleotide binds to $\beta$-tubulin the ribose hydroxyls are directed into the protein as a consequence of the reverse orientation of loop III (see table 2). That is, appropriate hydrogen bonding of the guanine ring to loop III requires that GTP rotate 180° about its long axis relative to its orientation in EF-Tu. At present it is not clear from experimental data, which indicate steric constraints around the 3′-hydroxyl, whether ribose hydroxyls are directed towards the surface or interior of tubulin (cf. [47]).

Although $\alpha$-tubulin has regions homologous to loops I, $I_A$, II and IV, it lacks sequence homology to loop III, the guanine-binding region. Whether this difference is sufficient to account for the non-exchangeable nature of the second GTP molecule in tubulin is unclear. Nevertheless, it suggests that the GTP-binding domain in $\alpha$-tubulin is conformationally different from that in $\beta$-tubulin.

Our model, like the Mandelkow 'model' [7], assumes that $\beta$-tubulin is an $\alpha/\beta$ protein. However, our model differs from this earlier model in several important respects: (i) The Mandelkow model implicates only one loop (loop $I_A$) in phosphoryl binding. Since ligand-binding sites in $\alpha/\beta$ proteins are generally formed from two adjacent loops [20], the Mandelkow model appears to be incomplete and is incapable in its present form of explaining differences in GTP and GDP binding to $\beta$-tubulin. (ii) In the Mandelkow model the guanine-binding site is formed from regions 60–69 and 240–244, and is a composite of sequences (region 60–69) observed in GTP-binding proteins such as EF-Tu [18] and sequences (region 240–244) observed in ATP/ADP- and dinucleotide-binding proteins [25]. However, these regions lack a sequence thought to confer guanine-binding specificity on the G-proteins [21,23]. Thus, the basis for the preferred binding of GNPs over ANPS in $\beta$-tubulin is unclear from the Mandelkow model.

Our analysis reveals that tubulin shares sequence homology with other GTP-binding proteins and meets the functional criteria for a G-protein [17]. In addition, tubulin is a substrate for ADP-ribosylation by cholera toxin [39], as are other G-proteins. However, tubulin also shares homology via sequence $I_A$ with ATP/ADP-binding proteins. A conclusive assignment of $\beta$-tubulin to a subclass of G-proteins appears to rest with 'loop' III which has the consensus sequence for guanine-binding specificity. If this sequence in $\beta$-tubulin is not essential for guanine binding then tubulin may be a highly divergent G-protein or alternative models which assign tubulin to a subclass of ATP/ADP proteins [7,46] may be more appropriate. We anticipate that recent advances in molecular biology of tubulin will help clarify the role of region III in GTP binding.

## ACKNOWLEDGEMENTS

We wish to thank W. Merrick for many stimulating discussions concerning G-proteins. We also would like to acknowledge our gratitude to M. Glynias for his assistance with secondary structure

analysis of tubulin. This work was supported in part by American Cancer Society Grant CD-228G to H.S.

REFERENCES

[1] Luduena, R., Shooter, E.M. and Wilson, L. (1977) J. Biol. Chem. 252, 7006–7013.
[2] Ponstingl, H., Krauhs, E., Little, M. and Kempf, T. (1981) Proc. Natl. Acad. Sci. USA 78, 2757–2761.
[3] Krauhs, E., Little, M., Kempf, T., Hofer-Warbinek, R., Ade, W. and Ponstingl, H. (1981) Proc. Natl. Acad. Sci. USA 78, 4156–4160.
[4] Valenzuela, P., Qoriga, M., Zaldivar, J., Rutter, W., Kirschner, M. and Cleveland, D. (1981) Nature 289, 650–655.
[5] Ponstingl, H., Krauhs, E., Little, M., Kempf, T., Hofer-Warbinek, R. and Ade, W. (1981B) Cold Spring Harbor Symp. Quant. Biol. 46, 191–197.
[6] Maccioni, R.B., Serrano, L. and Avila, J. (1985) BioEssays 2, 165–169.
[7] Mandelkow, E.M., Herrmann, M. and Ruhl, U. (1985) J. Mol. Biol. 185, 311–327.
[8] Geahlen, R.L. and Haley, B.E. (1979) J. Biol. Chem. 254, 11982–11987.
[9] Nath, J.P., Eagle, G.R. and Himes, R.H. (1985) Biochemistry 24, 1555–1560.
[10] Jemiolo, D.K. and Grishman, C.M. (1982) J. Biol. Chem. 257, 8148–8152.
[11] Huang, A.B., Lin, C.M. and Hamel, E. (1985) Biochim. Biophys. Acta 832, 22–32.
[12] Karr, T.L. and Purich, D.L. (1978) Biochem. Biophys. Res. Commun. 84, 957–961.
[13] Howard, W.D. and Timasheff, S.N. (1985) Fed. Proc. 44, 6082a.
[14] Morgan, J.L. (1985) Fed. Proc. 44, 1931a.
[15] Terry, B.J. and Purich, D.L. (1982) Adv. Enzymol. Related Areas Mol. Biol. 53, 113–161.
[16] Kirschner, M. and Mitchison, T. (1986) Cell 45, 329–342.
[17] Hughes, S.M. (1983) FEBS Lett. 164, 1–8.
[18] Leberman, R. and Egner, U. (1984) EMBO J. 4, 51–56.
[19] La Cour, T.F.M., Nyborg, J., Thirup, S. and Clark, B.F.C. (1985) EMBO J. 4, 2385–2388.
[20] Branden, C.I. (1980) Q. Rev. Biophys. 13, 318–347.
[21] McCormick, F., Clark, B.F.C., La Cour, T.F.M., Kjeldgaard, M., Norskov-Lauritsen, L. and Nyborg, J. (1985) Science 230, 78–82.
[22] Bourne, H.R. (1986) Nature 321, 814–816.
[23] Dever, T.E., Glynias, M.J. and Merrick, W.C. (1987) Proc. Natl. Acad. Sci. USA, in press.
[24] Little, M., Krauhs, E. and Ponstingl, H. (1981) BioSystems 14, 239–246.
[25] Sternberg, M.J.E. and Taylor, W.R. (1984) FEBS Lett. 175, 387–390.
[26] Wierenga, R.K., Terpstra, P. and Hol, W.G.J. (1986) J. Mol. Biol. 187, 101–107.
[27] Lee, J.C., Corfman, D., Frigon, R.P. and Timasheff, S.N. (1978) Arch. Biochem. Biophys. 185, 4–14.
[28] Correia, J.J., Bary, L.T. and Williams, R.C. jr (1968) Symposium on Molecular Biology of the Cytoskeleton, Airlie House, Airlie, VA.
[29] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[30] Blank, G.S., Yaffe, M.B., Szasz, J., George, E., Rosenberry, T.L and Sternlicht, H. (1986) Ann. NY Acad. Sci. 466, 467–481.
[31] Leszczynski, J.F. and Rose, G.D. (1986) Science 234, 849–855.
[32] Garnier, J., Osguthrope, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[33] Chou, P.Y. and Fasman, G.D. (1978) Annu. Rev. Biochem. 47, 251–276.
[34] Miller, D.L. and Weissbach, H. (1970) Arch. Biochem. Biophys. 141, 26–37.
[35] Olmsted, J.B. and Borisy, G.G. (1975) Biochemistry 14, 2996–3005.
[36] Bayley, P.M. and Manser, E.J. (1985) Nature 318, 683–685.
[37] Nath, J.P., Eagle, G.R. and Himes, R.H. (1986) Ann. NY Acad. Sci. 466, 482–495.
[38] Sullivan, K.F. and Cleveland, D.W. (1986) Proc. Natl. Acad. Sci. USA 83, 4327–4331.
[39] Hawkins, D.J. and Browning, E.T. (1982) Biochemistry 21, 4474–4479.
[40] Richardson, J.S. (1981) Adv. Protein Chem. 34, 167–339.
[41] Richardson, J.S. (1976) Proc. Natl. Acad. Sci. USA 73, 2619–2623.
[42] Cowan, N.J., Dobner, P., Fuchs, E.V. and Cleveland, D.W. (1983) Mol. Cell Biol. 3, 1738–1745.
[43] Pai, E.F., Sachsenmaier, W., Schirmer, R.H. and Schulz, G.E. (1977) J. Mol. Biol. 114, 37–45.
[44] Frohlich, K.U., Entian, K.D. and Mecke, D. (1985) Gene 36, 105–111.
[45] Shoham, M. and Steitz, T.A. (1980) J. Mol. Biol. 140, 1–14.
[46] Linse, K. and Mandelkow, E.M. (1986) J. Cell Biol. 103, 545a.
[47] Hamel, E. and Lin, C.M. (1984) Biochim. Biophys. Acta 797, 117–127.